Automated Medical Knowledge Acquisition: A Study of Consistency

Gwendolyn C. Murphy, Ph.D. and Charles P. Friedman, Ph.D.
University of North Carolina at Chapel Hill

ABSTRACT

Knowledge bases are more representative of the population of medical experts if they are constructed by a group of individuals, rather than one practitioner. However, one runs into problems with consistency when information is elicited from a group without a consistent format and terminology. This study examines the consistency of relatively unconstrained computer-elicited medical knowledge using the computer program, KSS0. The results of this study show that the group of ten general internists were somewhat consistent in the diagnoses they listed for a patient presenting with chest pain. They were much less consistent in the findings they listed to differentiate between the diagnoses they had listed. The mean number of subjects listing each diagnosis was 3.3 $\pm$2.7 while the mean for findings was 2.0 $\pm$ 1.5. The implications of these data are discussed.

INTRODUCTION

Until recently, development and maintenance of many knowledge bases (KBs) has been a task that required the services of a knowledge engineer to translate the elicited data into machine language [1]. Most knowledge bases have been constructed by having one expert develop one disease profile at a time. The disease profile is then evaluated and modified by a group of experts [2]. Because it has been so time-consuming a task, only a small group of people has usually been involved. Rarely are multiple experts asked to develop a profile for the same disease [3]. Consequently, the acquired knowledge may reflect representations particular to the small group of authors. To increase the generalizability of the KB, it could be based on the collaboration of many different groups of individuals. Automated knowledge acquisition is now making this possible.

Several computer programs have been developed to facilitate knowledge acquisition by different individuals working on the same project [1,4, 5, 6]. The validity and reproducibility of such an approach is beginning to be documented [3]. One program that was designed to elicit knowledge from individuals in an unconstrained way and without the need for a knowledge engineer is Knowledge Support System Zero (KSS0). KSS0 is an "implicit", problem-solving program because the user does not need to understand how the program works [7]. A program of this type provides a method to build a representation of the subject's conceptual structure without direct elicitation. Instead, the subject provides examples within the domain of interest and then states in concrete terms how to distinguish between the examples [8]. Shaw and Gaines have developed the program KSS0 and have used it to elicit and compare information from multiple experts [9]. According to Shaw, hierarchical and spatial cluster analysis of the gathered data can be used to develop the conceptual structure for that subject [10].

KSS0 is based on personal construct psychology of Kelly [11]. Kelly proposed that individuals use their experience to develop constructs, which are defined as abstract qualities that one uses to model reality. As each person tests these constructs against reality, s/he revises them. Kelly hypothesized that each person evolves a finite number of constructs and attributes that go along with those constructs. The constructs and attributes are related in such a way that a grid, called a repertory grid, can be produced from the elicited data and an hierarchical classification can be developed from the grid.

A major problem that has been encountered with multi-center knowledge acquisition is consistency of the product. Medicine is such a complex field that there are usually several possible ways to approach any problem. As a result, it is difficult to find an agreed upon standard. Instead, standards of care vary from region to region, and sometimes even from practitioner to practitioner. This study answers the question of how consistent is a group of general internists in how they describe the diagnosis of a patient who presents with chest pain. Specifically, what diagnoses are on their differential, and what findings (history, physical, and laboratory) do they use to distinguish between the diagnoses?

PROCEDURE

The subjects were ten relatively young,

0195-4210/94/$5.00 © 1994 AMIA, Inc.

faculty-level physicians, each of whom had finished three years of residency training and was board certified in internal medicine. They had been in practice an average of 5.3 years (range 1-11 years). These physicians were all participants in a Faculty Development Fellowship in General Internal Medicine at the University of North Carolina at Chapel Hill, but they came from ten different academic institutions from across the mid-Atlantic region. Eight of the ten subjects estimated they had encountered one hundred to one thousand patients presenting with chest pain. Two of the subjects estimated they had seen over a thousand such patients.

The program, KSSO, asks the subject to generate a list of constructs (diagnoses in this study) and attributes (findings in this study). The diagnoses and findings were to be applicable to a patient presenting with chest pain. The subject then entered the list of diagnoses into the computer. For this portion, subjects were limited to a maximum of 15 diagnoses because that was the most that could fit on the computer screen simultaneously. In some cases the subjects had to reduce their initial list of diagnoses. They were told to choose the most likely and/or important diagnoses.

Next, the subject was directed to enter a finding (history, physical exam, or laboratory test) by specifying the result and its opposite (e.g., febrile--afebrile). The computer positioned the finding and its opposite at either end of a line. The subject then placed each of the previously listed diagnoses where it seemed to belong on a line between the finding and its opposite. For example, if the finding was febrile-afebrile, pneumonia would likely be at the febrile pole, and angina would likely be at the afebrile pole. See Figure 1.

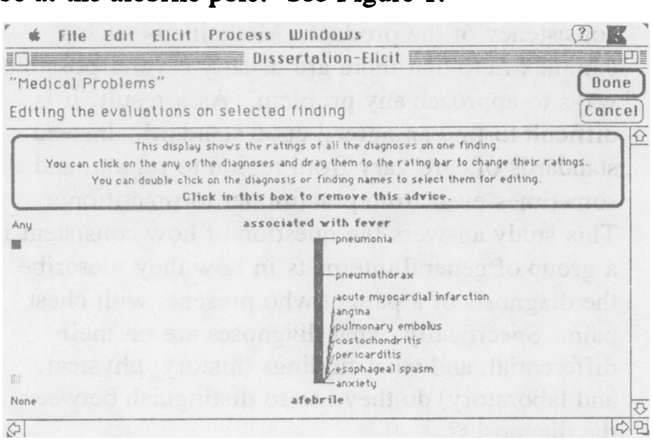

Figure 1
Rating the diagnoses on one finding

The subject placed all the diagnoses on the scale where s/he thought they belonged. Once the subject had moved all the diagnoses to their desired position, the subject could proceed to the next screen for the next finding. This process continued until the subject had entered all the findings s/he could. For each finding, the program recorded a number from 1-9 to denote the position of the diagnosis on the finding. KSSO used these values to construct a repertory grid of the diagnoses by findings. See Figure 2.

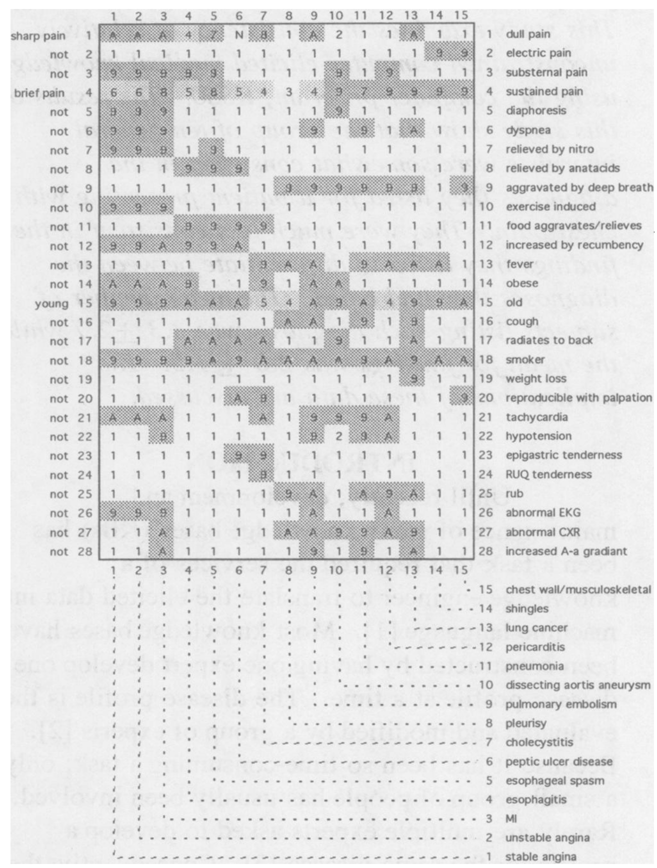

Figure 2
Example of a repertory grid

The repertory grids provide a large amount of data. Terms that were synonymous (such as shortness of breath and tachypnea) were merged by the first author and verified by an experienced family physician. Any question of whether they were synonymous resulted in them not being merged. The diagnoses and findings were sorted by the number of subjects listing each. The mean and standard deviation was calculated for the number of subjects listing each diagnosis or finding. The degree of overlap in diagnoses or findings can be summarized by a coefficient of consistency (CC) in which the number of elements listed by two or more subjects is divided by the number of entries listed after merging synonymous terms.

$$CC = \frac{E_c}{E_m}$$

CC=coefficient of consistency
Ec= the number of consistent elements listed
Em=number of merged elements

This coefficient can vary between one (if all of the subjects list the same elements) and zero if there are no elements in common. The results are presented below.

RESULTS

The data from this study are based on a single session with each subject. The terms used by the subjects to express diagnoses varied between general terms and very specific ones. For example, for musculoskeletal diseases, some subjects listed "chest wall--musculoskeletal" while others listed "rib fracture" or "costochondritis". All or nearly all of the subjects had one or more diagnoses in the areas of musculoskeletal, cardiac, pulmonary, and gastrointestinal disease. Several diagnoses were listed by seven or more subjects, including angina, pneumonia, pulmonary embolus, chest wall/musculoskeletal, myocardial infarct, pericarditis, and pleurisy. There were 14 diagnoses listed by only one subject. The findings listed showed much more variability than did the diagnoses. There were 78 findings listed by only one subject, and only two findings listed by seven or more subjects. The mean number of physicians listing each diagnosis and finding was 3.3 and 2.0 respectively. The coefficient of consistency for the diagnoses and findings was .60 and .40 respectively. See Table 1 below for the means, standard deviations, and coefficients of consistency.

Table 1
Means, standard deviations and coefficients of consistency for diagnoses and findings

Type of Entry	# MDs/entry mean ± s.d.	CC
Diagnoses	3.3 ± 2.7	.60
Findings	2.0 ± 1.5	.40

The frequencies of diagnoses and findings are shown by Figures 3 and 4.

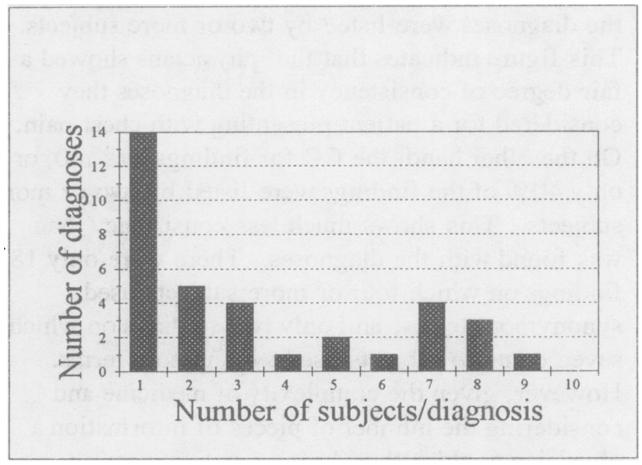

Figure 3: Frequencies of diagnoses listed
Number of subjects listing the same or a synonymous diagnosis using KSSO

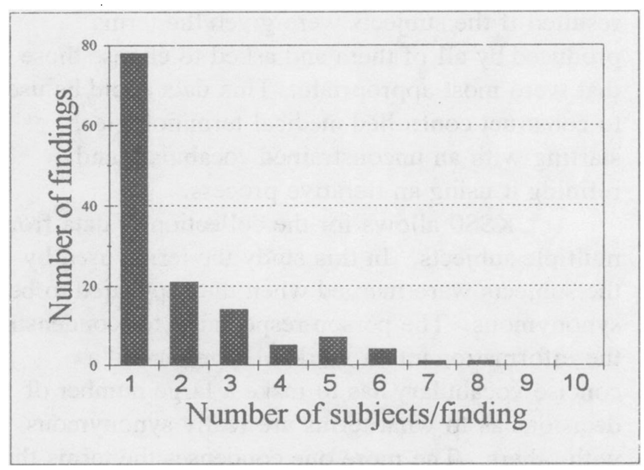

Figure 4: Frequencies of findings listed
Number of subjects listing the same or a synonymous finding using the program KSSO

DISCUSSION

KSSO is a knowledge acquisition tool that offers a relatively unconstrained graphical interface with which the user interacts directly. The question being answered in this study was how consistent a group of general internists was in the diagnoses and findings they used to describe the diagnosis of a patient who presented with chest pain. The results show that when there are few constraints put on the format or terminology to be used, the subjects were somewhat consistent in the diagnoses they listed, and much less consistent in the findings they listed.

A coefficient of consistency (CC) is defined and calculated. It can be interpreted as the percentage of the entries listed by two or more subjects. The CC for diagnoses was .60, or 60% of the diagnoses were listed by two or more subjects. This figure indicates that the physicians showed a fair degree of consistency in the diagnoses they considered for a patient presenting with chest pain. On the other hand, the CC for findings was .40, or only 40% of the findings were listed by two or more subjects. This shows much less consistency than was found with the diagnoses. There were only 18 findings on which four or more subjects used synonymous terms, and only two findings on which seven or more subjects used synonymous terms. However, given the complexity of medicine and considering the number of pieces of information a physician could gather during a patient session, perhaps a coefficient of .40 shows reasonable consistency. It must also be remembered that the subjects were allowed only one iteration of data elicitation. More consistency would probably have resulted if the subjects were given the terms produced by all of them and asked to choose those that were most appropriate. This data could be used to construct controlled medical terminologies, starting with an unconstrained vocabulary and refining it using an iterative process.

KSSO allows for the collection of data from multiple subjects. In this study the terms used by the subjects were merged when they appeared to be synonymous. The person responsible for condensing the information into a workable format and a concise vocabulary has to make a large number of decisions as to what terms are really synonymous with others. The more one condenses the terms the higher the degree of consistency that is achieved. Effort was made to merge only those terms that were clearly synonymous. But, because of the subjectivity that is possible in this step, it must be considered a weakness in the methods. A second weakness was the restriction of subjects to no more than 15 diagnoses. This would tend to increase the degree of consistency by forcing subjects to choose only the 15 most likely or important. The data collected might be further strengthened by allowing the subjects to choose the best diagnoses and findings from the merged list and to have them use KSSO a second time using their revised diagnoses and findings. Overall, the method used was a reasonable one, but could be strengthened.

CONCLUSIONS

Expert knowledge acquisition and representation are crucial elements in the production of an expert system knowledge base or a standardized protocol for use in clinical practice. The elicitation of inter-related medical information from an expert takes a great deal of time. Because of the time involved, knowledge is usually elicited from only one expert, with verification of that knowledge by a group of experts. However, a knowledge base would be more representative of medical expertise if it were constructed by a group of experts rather than by one expert.

There are problems with eliciting knowledge from multiple medical experts, not the least of which is inconsistency in the elicited information. Medical knowledge is not a stable and agreed upon set of facts. Instead, it is a rapidly changing body of knowledge that has regional and individual variation. In the past, the format and terminology have often been standardized prior to elicitation of knowledge [12,13,14]. This forces the subject to modify his/her understanding to fit the format. Cognitive research argues that in constraining the format, one changes the elicited knowledge to the point where it may no longer conform to the knowledge the subject generally uses [15]. Perhaps, rather than standardizing the format and terminology prior to knowledge elicitation, it would be more valid to elicit knowledge in a relatively unconstrained format and then determine the most appropriate format. The data presented here offer a first step toward that end.

Acknowledgements
We would like to thank Mark Musen M.D., Ph.D. for his guidance during the preparation of this paper.

References

1. Giuse, N.B., Giuse, D.A., & Miller, R.A. Computer assisted multi-center creation of medical knowledge bases. In *Proceedings of the Twelfth Annual Symposium on Computer Applications in Medical Care,* pages 583-90. IEEE Computer Society Press, Washington, D.C., 1988.

2. Giuse, N.B., Bankowitz, R.A., Giuse, D.A., Parker, R.C., Miller, R.A., Medical knowledge base acquisition: the role of the expert review process in disease profile construction. In *Proceedings of the Thirteenth Annual Symposium on Computer Applications in Medical Care,* pages 105-9, IEEE Computer Society Press, Washington, D.C., 1989.

3. Giuse, N.B. et. al., Evaluating consensus among physicians in medical knowledge base construction. *Methods of Information in Medicine* 32: 137-45, 1993.

4. Kingsland, L.C. III, Lindberg D.A.B. The criteria form of knowledge representation in medical artificial intelligence. MEDINFO 86 Proceedings pages 12-16, North-Holland, Amsterdam, 1986.

5. Miller, P.L., Blumenfrucht, S. J., Rose, J.R., et. al., Expert system knowledge acquisition for domains of medical workup: An augmented transition network model. In *Proceedings of the Tenth Annual Symposium on Computer Applications in Medical Care,* pages 30-35. IEEE Computer Society Press, Washington, D.C., 1986.

6. Weiss, S. M.,Politakis, P.G., & Ginsberg, A. Empirical analysis and refinement of expert system knowledge bases. In *Proceedings of the Tenth Annual Symposium on Computer Applications in Medical Care,* pages 53-60. IEEE Computer Society Press, Washington, D.C., 1986.

7. Musen, M.A .*Automated generation of model-based knowledge-acquisition tools*. San Mateo, CA: Morgan Kaufmann Publishers, Inc., 1989.

8. Shaw, M.L.G., & Gaines, B.R. KITTEN:Knowledge initiation and transfer tools for experts and novices. *International Journal of Man-Machine Studies* 27:251-280, 1987.

9. Shaw, M.L.G., & Gaines, B.R. Comparing conceptual structures: consensus, conflict, correspondence and contrast. *Knowledge Acquisition* 1:341-363, 1989.

10. Shaw, M.L.G. *On becoming a personal scientist: Interactive computer elicitation of personal models of the world.* London: Academic Press, 1980.

11. Kelly, G.A., *The psychology of personal constructs.* N.Y.:Norton, 1955.

12. Bouhaddou, O., Warner, H.R., Yu, H., Lincoln, M.J. The knowledge capabilities of the vocabulary component of a medical expert system. In Miller, R.A. (ed) *Proceedings of the Fourteenth Annual Symposium on Computer Applications in Medical Care,* pages 655-60, IEEE Computer Society Press, Washington, D.C., 1990.

13. Cimino, J.J. Representation of Clinical laboratory terminology in the unified medical language system. In *Proceedings of the Fifteenth Annual Symposium on Computer Applications in Medical Care,* pages 199-203, IEEE Computer Society Press, Washington, D.C., 1991.

14. Masarie, F.S.,Jr., Miller, R.A., Bouhaddou, O, Giuse, N. B., & Warner, H. R. An interlingua for electronic interchange of medical information: using frames to map between clinical vocabularies. *Computers and biomedical Research* 24:379-400, 1991.

15. Adair, J.G. & Spinner, B. Subject's access to cognitive processes: Demand characteristics and verbal report. *Journal of the theory of social behavior* 11:31-52, 1981.